Ahmed Metwally

Department of Computer Science, UCSB
Santa Barbara CA, 93106
(805) 403-9725
metwally@cs.ucsb.edu

Education:

2002 - 2007: PhD; UC Santa Barbara, Department of Computer Science.

2002 - 2006: MS; UC Santa Barbara, Department of Computer Science.

2000 - 2002: MS Candidate; Alexandria University, Faculty of Engineering, Computer and Systems Engineering Department. Degree unfinished.

1995 - 2000: BS; Alexandria University, Faculty of Engineering, Computer and Systems Engineering Department.

Dissertation Topic:

My dissertation entitled "*On-Line Data Forensics for Fraud Detection in Internet Advertising*" proposes a new approach for detecting fraud in Internet advertising through analyzing traffic logs. Due to the scale of the data, new algorithms that analyze the traffic using limited time and space are proposed to discover traffic abnormalities that can uncover fraudulent traffic. The algorithms provide very tight guarantees on errors. When the proposed algorithms were tested on real data, numerous fraud instances were discovered.

Publications:

Journal Articles and Book Chapters:

- [M09] **Ahmed Metwally**: "*Frequent Items on Streams*". Springer Publishers' **Encyclopedia of Database Systems**, Tamer Özsu and Ling Liu (Eds.), pages: 1175-1179, 2009.
- [MEAE08] **Ahmed Metwally**, Fatih Emekci, Divyakant Agrawal, and Amr El Abbadi: "*SLEUTH: Single-pubLisher attack dEtection Using correlaTion Hunting*". **PVLDB** Proceedings of the Very Large Data Bases Endowment, Vol. 1, No. 2, pages: 1217-1228, 2008.
- [MAE06] **Ahmed Metwally**, Divyakant Agrawal, and Amr El Abbadi: "*An Integrated Efficient Solution for Computing Frequent and Top-k Elements in Data Streams*". ACM **TODS** Transactions On Database Systems, Vol. 31, No. 3, pages: 1095-1133, September 2006.

Conference Papers:

- [KTPMDCB09] Carmelo Kintana, David Turner, Jia-Yu Pan, **Ahmed Metwally,** Neil Daswani, Erika Chin, and Andrew Bortz: "*The Goals and Challenges of Click Fraud Penetration Testing Systems*". IEEE **ISSRE** International Symposium on Software Reliability Engineering, 2008, to appear.

- [AEEM09] Divyakant Agrawal, Amr El Abbadi, Fatih Emekci, and **Ahmed Metwally**: "*Database Management as a Service: Challenges and Opportunities*". IEEE **ICDE** International Conference on Data Engineering, pages: 1709-1716, 2009.
- [MAE08] **Ahmed Metwally**, Divyakant Agrawal, Amr El Abbadi: "*Why Go Logarithmic if We Can Go Linear? Towards Effective Distinct Counting of Search Traffic*". **EDBT** International Conference on Extending Database Technology, pages: 618-629, 2008.
- [MAEZ07] **Ahmed Metwally**, Divyakant Agrawal, Amr El Abbadi, and Qi Zheng: "*On Hit Inflation Techniques and Detection in Streams of Web Advertising Networks*". IEEE **ICDCS** International Conference on Distributed Computing Systems, paper number 52, 2007.
- [BMAE07a] Nagender Bandi, **Ahmed Metwally**, Divyakant Agrawal, and Amr El Abbadi: "*Fast Algorithms for Data Streams using Associative Memories*", ACM **SIGMOD** International Conference on Management of Data, pages: 247-256, 2007.
- [BMAE07b] Nagender Bandi, **Ahmed Metwally**, Divyakant Agrawal, and Amr El Abbadi: "*TCAM-conscious Algorithms for Data Streams*", IEEE **ICDE** International Conference on Data Engineering, pages: 1342-1344, 2007.
- [MAE07] **Ahmed Metwally**, Divyakant Agrawal, and Amr El Abbadi: "*DETECTIVES: DETEcting Coalition hiT Inflation attacks in adVertising nEtworks Streams*". The International **WWW** Conference, pages: 241-250, 2007.
- [MAE05a] **Ahmed Metwally**, Divyakant Agrawal, and Amr El Abbadi: "*Using Association Rules for Fraud Detection in Web Advertising Networks*". **VLDB** International Conference on Very Large Data Bases, pages: 169-180, 2005.
- [MAE05b] **Ahmed Metwally**, Divyakant Agrawal, and Amr El Abbadi: "*Duplicate Detection in Click Streams*". The International **WWW** Conference, pages: 12-21, 2005.
- [MAE05c] **Ahmed Metwally**, Divyakant Agrawal, and Amr El Abbadi: "*Efficient Computation of Frequent and Top-k Elements in Data Streams*". **ICDT** International Conference on Database Theory, pages: 398-412, 2005.
- [FAEM04] Ying Feng, Divyakant Agrawal, Amr El Abbadi, and **Ahmed Metwally**: "*Range CUBE: Efficient Cube Computation by Exploiting Data Correlation*". IEEE **ICDE** International Conference on Data Engineering, pages: 658-670, 2004.

Ongoing Research:

- [MAEE] **Ahmed Metwally**, Divyakant Agrawal, Amr El Abbadi, and Fatih Emekci: "*Why Go Logarithmic if We Can Go Linear? Reviving Linear Counting for Stream Applications*". Journal Manuscript under preparation.
- [MAEZ] **Ahmed Metwally**, Divyakant Agrawal, Amr El Abbadi, and Qi Zheng: "*Hide and Seek: Detecting Click Spam in Streams of Web Advertising Networks*". Manuscript under preparation.
- [EMAEa] Fatih Emekci, **Ahmed Metwally**, Divyakant Agrawal, and Amr El Abbadi: "*A Privacy Preserving Data Outsourcing Framework*". Journal manuscript under preparation.
- [EMAEb] Fatih Emekci, **Ahmed Metwally**, Divyakant Agrawal, and Amr El Abbadi: "*ABACUS: A Distributed Middleware For Privacy Preserving Data Sharing Across Private Data*". Journal manuscript under preparation.

Academic Services:

Program Committee Member: SIGMOD 2010.

Journal Reviewer: ACM TODS Transactions On Database Systems.
ACM TOSN Transactions On Sensor Networks.
CACM Communications of the ACM.
IEEE Communications Letters.

External Reviewer: with Prof. D. Agrawal, and Prof. A. El Abbadi, SIGMOD, VLDB, PODS, ICDT, ICDE, ACM-GIS, EDBT, ICDCS.

Mentorship: Mentored intern Amr Ebaid in the Ad Traffic Quality Team at Google.
Mentored the "Vacation Recommendation" graduation project at Alexandria University, 2008.

Teaching Assistant:

- **Fall 2002:** UC Santa Barbara, Department of Computer Science, CS160 "*Compiler Design*".
- **Spring 2002:** UC Santa Barbara, Department of Computer Science, CS60 "*Unix/Linux/C/C++*".
- **Fall 2003:** UC Santa Barbara, Department of Computer Science, CS40 "*Discrete Mathematics*".

University Committee Membership:

- Computer Science Association of Graduate Students, representative of faculty recruiting, 2006/2007.
- Secretary of Computer Science Association of Graduate Students, 2006/2007.
- Computer Science Graduate Student Research Seminar, Organization and Reviewing, 2006/2007.
- UCSB Graduate Students Association Committee Representatives, GSA Excellence in Teaching Awards, 2005/2006.

- Computer Science Association of Graduate Students, representative of curriculum affairs, 2005/2006.

Honors and Awards:

Winter 2003 - Spring 2003: Full scholarship awarded by Commission Junction, A ValueClick Company.

2002 -2008: Full scholarship awarded by University of California, Santa Barbara for Ph.D. study (2002-2008).

2003 -2007: Presidential Work Study Research Grant.

1995 -2000: High-Honor computer engineering student in all the semesters of 1995-2000.

Employment History:

April 2008 - Now: Software Engineer at Google, AdSpam (Ad Traffic Quality) Team.

Responsibilities:

1- Research and develop methods for signals in traffic that reveal fraudulent activities by analyzing massive traffic logs.
   a. Estimating the number of users behind an IP of a specific node in a multidimensional cube (e.g. performing certain activities, appearing in a specific time range, and belonging to a specific country).
   b. Analyzing the behavior of the normal users behind an IP.
   c. Devising new algorithms and metrics to detect abnormal activities given the IP profile.

2- Tune new and existing systems and processes that identify click spam based on experimentation with Google data.

3- Enhance the existing Ad Spam architecture by implementing libraries of approximate stream analysis algorithms for faster analysis.

June 2007 - April 2008: Software Engineer at Oracle, Query Optimization Group.

Responsibilities:

1- Building a random generator of database workloads, including schemas, tables, and queries. I am responsible for the random query generation given a schema DDL. The objectives are as follows:
   d. The frequency of generating the various constructs of the DML language could be dynamically changed at run time. For each generated query, the framework examines which parts of Oracle's sever is being tested. The server subroutines (e.g. query transformations) that are not frequently tested using the generated queries should feedback into the probability knobs that generate the DML constructs. These knobs should automatically adapt and generate queries with constructs that test the non-tested server subroutines.
   e. The semantic context of the query should be discoverable by traversing the parse tree. Hence, any part of the query can be altered while not violating the semantic correctness of the query. The goal is to

grow the complexities of the generated queries gradually for easier debugging of queries causing the server to crash or producing wrong results

October 2006 - March 2007: Internship at Ask.com.

Responsibilities:

1- Designing and implementing a new approximate algorithm (in C++) for counting the number of distinct elements in a huge dataset using only one scan on the data and limited space. Doing the statistical analysis to prove the error guarantees. Implementing all the existing algorithms in the stream distinct counting literature to compare the experimental results on real data.

2- Analyzing customer behavior, and designing a metric to measure customer retention.

3- Assisting Divyakant Agrawal (Vice President of Warehousing) in designing the main warehouse of click streams.

Jun 2006 - Sep 2006: Internship at Google Inc., Infrastructure Department, Log Analysis Team.

Responsibilities:

1- Designing and implementing a low-level C++ library for stream algorithms. Designing and implementing algorithms to find approximate frequent elements and exact Min/Max-k elements in amortized constant time per stream element. Modifying already known algorithms for discovering distinct count and sum over variable-sized sliding window.

Apr 2005 - Jan 2006: Internship at Fastclick; Researcher, responsible for the fraud detection project, and involved in the advertisement optimization project and query optimization process.

Responsibilities:

1- Rebuilding the fraud detection system (written in Perl and interfacing with Mysql database) to make more scalable and to add the feature of making the analysis based on a sliding window of the traffic. This entailed redesigning of the entire program and database schema used.

2- Devising new algorithms for fraud detection by doing traffic stream analysis (C++). Numerous fraudsters were detected using those algorithms.

3- Administrating the Spam servers (Perl).

4- Proposing new methods for advertisement optimization. The methods included techniques to increase the fill-ratio, and considering the remaining daily budgets of advertisers, in addition to changing the advertisement pick engine to do exploration to optimize which advertisements work best on which sites.

5- Optimizing frequent SQL query through adding and /or dropping indices, creating views, etc.

Jan 2004 - Jun 2004: Internship at Commission Junction, A ValueClick Company; Cognos ReportNet Developer; and Designer of the Ad Optimization Project.

Responsibilities:

1- Working with Per Pettersen (CTO) to devise an algorithm for advertisement optimization based on the theory of exploration versus exploitation, and producing a complete design for the project.
2- Using the Cognos ReportNet API to build a Java portlet that interfaces with the ReportNet server.

Jun 2003 - Sep 2003: Internship at Commission Junction, A ValueClick Company; Warehouse Designer and Developer.

Responsibilities:

4- Attending Informatica PowerCenter formal training.
5- Being one of the team who built and administrated a data-warehousing engine using PowerCenter.

Dec 2001 - Jun 2002: Bibliotheca Alexandrina, Alexandria. Egypt; Programmer/IT Engineer.

Responsibilities:

1- Administrating the clients and server of the VTLS software.
2- Designing and writing a simplified manual for MS Office for the librarians.

Jul 2000 - Dec 2001: Advanced Computer technology (ACT); Programmer/Analyst of ERP systems.

Responsibilities:

1- Programmer of the ERP systems for the modules of Fixed Assets, Accounts Payable, Inventory Management.
2- Report developer using Crystal Reports for the corresponding ERP reports.